

Cien años de robótica en la Ciencia Ficción

Campoamor Stursberg, Rutwig

Para el aficionado incondicional a la ciencia ficción, el año 1920 destaca especialmente entre las efemérides asociadas al género, al marcar el centenario del nacimiento de Isaac Asimov, una de las figuras que han tenido mayor repercusión, difusión e influencia en la literatura de ciencia ficción.¹ En este contexto, el nombre de Asimov se asocia de forma natural con una de las corrientes más características de la literatura de anticipación, los robots. Aunque la idea de entes mecánicos o mecanizados es anterior a Asimov y a la propia ciencia ficción,² le corresponde a él el mérito de haber conferido a la noción de robot una relevancia literaria y técnica propias, a la vez de intentar sistematizar e incluso proponer pautas de comportamiento que, en un futuro indeterminado, pudiesen servir como base de una programación eficiente y segura de máquinas autónomas dotadas de una capacidad de aprendizaje y asimilación. El desarrollo del "cerebro positrónico" es la materialización física del deseo de un intelecto artificial con todas las características creativas e intelectuales humanas, pero desprovisto de sus vicios y defectos, que son rigurosamente descartados mediante la aplicación de tres leyes, que forman el decálogo de la psicología robótica.

Aunque las tres célebres leyes de la robótica sean criticables desde una perspectiva puramente científica, al constituir postulados de naturaleza ética y/o moral en lugar de axiomas o proposiciones pertenecientes a la lógica, es innegable que su formulación ha sido de gran utilidad tanto en el contexto de la teoría de la información como en la propia robótica, aunque sea en términos meramente comparativos, sirviendo a menudo como motivación para los especialistas. Recordemos que el industrial Joseph Engelberger, fascinado por las obras de Asimov, desarrollaría el primer robot industrial (llamado Unimate #001) en 1958.³ Por otra parte, es muy probable que las

peripecias y tribulaciones que padecen los robots asimovianos hayan sido de algún modo una inspiración para los especialistas en computación, con el fin de programar y emular pautas del comportamiento y la psicología humanas, en lo que puede interpretarse como un antecedente de la psicología robótica.⁴ El término robot como tal aparece incontestablemente por primera vez en una pieza teatral en tres actos titulada R.U.R. (Rossumovi Univerzální Roboti), debida a Karel Čapek, y fechada en 1920.⁵ Puede considerarse, por tanto, que dicho año conmemora tanto el nacimiento de la robótica moderna como el de su principal impulsor en la ciencia ficción.

Los robots de Čapek no son meramente ingenios mecánicos con apariencia humana, sino entes orgánicos antropomorfos producidos, eso sí, mediante un proceso industrial. En esencia, podrían considerarse como androides producidos en serie. La elección de su denominación como "robot", vocablo derivado del checo "robota" y que significa trabajo (forzado), se debe a su función principal, que es liberar a la humanidad del trabajo para ofrecerle el retorno al paraíso. Dicha liberación equivale, no obstante, a la sumisión y esclavitud de los robots, que en un principio son indiferentes a su destino. La llegada de la protagonista Helena Glory a la remota isla donde se encuentran las industrias de R.U.R. tendrá funestas consecuencias. Romántica e inconsciente activista en su lucha por la equiparación moral, legal y espiritual de los robots, su matrimonio con el industrial Domin la pondrá en una situación privilegiada para influir en el trabajo del Dr. Gall, fisiólogo jefe de la empresa. A instancias de la bienintencionada Helena, Gall tratará de dotar a sus robots de un "alma", y los experimentos para modificar la psicología de los robots darán resultado, aunque no exactamente en el sentido deseado. Al desarrollar su conciencia y una especie de moral, los robots

se percatan de su esclavitud y aprenden a odiar a sus opresores, a los que acusan de privarles del "secreto de la vida". Alentados y capitaneados por el robot Radius, los androides se sublevan mundialmente contra la humanidad y la exterminan, con la excepción del constructor Alquist, a quien los robots exigen, piden y finalmente suplican el citado secreto de la vida, que no es sino el manuscrito de Rossum que permitió la creación de los androides. Habiendo sido destruido este valioso documento por la propia Helena en un arrebato de pánico, los robots deben asumir que están, a su vez, condenados a la extinción.

En su conjunto, la obra es, como otras tantas de Čapek, pronunciadamente pesimista, aunque a la hora de interpretar R.U.R. no debe olvidarse el contexto histórico en el cual la obra fue escrita, recientemente acabada una guerra donde los progresos científicos habían sido inmediatamente adaptados al uso bélico. Un error ampliamente extendido fue considerar la obra de Čapek como anticientífica, del cual se derivó la imagen del robot como un ente hostil hacia el ser humano, un tópico que fue rápidamente asimilado y burdamente identificado con el llamado "complejo de Frankenstein" por multitud de autores e historietistas de la ciencia ficción de producción en los años treinta y cuarenta. Fiel a su filosofía declaradamente humanista, Čapek deplora que el avance científico no tenga en cuenta las posibles consecuencias sociales y morales, pero en ningún momento condena la ciencia en sí misma.

Al margen de esta obra teatral, que marca el punto de partida de la noción moderna de robot en la ciencia ficción, y que posteriormente trascenderá para ser adoptada en la industria y la técnica, hay varios antecedentes en la literatura fantástica que anticipan la era de la robótica, si bien bajo una terminología distinta, donde tales ingenios son denominados de forma cáustica "hombres mecánicos" o autómatas, cuando no son abiertamente declaradas como monstruosidades o sacrílegas creaciones. En este marco serían clasificables *L'Eve future* (1886) de Villiers de l'Isle-Adam, en la que un inventor construye un autómata para

tratar de ayudar a un buen amigo sumido en la desesperación por un amor no correspondido, así como *Ignis* (1883) de Didier de Chousy, una de las primeras descripciones de una civilización de tipo robótico, en la que incluso se nos describe una revolución de la "plebe de metal". En la novela *Metroplis* (1926) de Thea von Harbou, la figura del robot también está concebida como una herramienta de opresión, en este caso, para agitar a las masas y posibilitar una intervención de castigo por parte de la autocracia.

La notable excepción a esta regla viene dada por la novela *Utazás Faremido* del escritor húngaro Fryges Karinthy, publicada en 1916 y concebida como una fantasía de tipo Gulliver ambientada en la Primera Guerra Mundial. En ella, el protagonista de la historia es rescatado de un naufragio por unos extraños seres inorgánicos procedentes del planeta Faremido. Aunque no se menciona explícitamente el vocablo "robot", los habitantes de Faremido son claramente de tipo robótico. Su principal medio de comunicación es una lengua de tipo musical (de ahí el título del libro), que enseñan al protagonista, además de tratar de convencerle de la superioridad de la vida inorgánica sobre aquella basada en el carbono, que se considera una enfermedad. La novela contiene una amplia e interesante descripción del proceso de fabricación de estos curiosos alienígenas, así como un ingrediente revolucionario digno de una obra de Dick. Mediante una droga evidentemente sintética, el héroe de la historia puede vislumbrar y experimentar brevemente las sensaciones y motivaciones existenciales de los habitantes de Faremido, antes de que éstos reexpidan al héroe a la Tierra, donde presumiblemente el recuerdo de las magnificencias vividas atormenten la existencia del protagonista.

Curiosamente, el inconfundible título *I, Robot* no se debe a Asimov y a su célebre compilación de las historias de robots aparecidas en la revista *Astounding*,⁶ sino a una serie de relatos debidos a la pluma de Eando Binder, un autor prolífico en la década de 1930-1940 y hoy prácticamente olvidado, en la que un robot llamado Adam

Link cuenta sus diversas aventuras. Las siete historias que componen el ciclo de Adam Link, publicadas en *Amazing Stories* entre 1939 y 1942 y recopiladas en forma de libro en 1965, son ciertamente pueriles en su argumento y desarrollo, aunque son meritorias por ser de las primeras en las que un robot exhibe principios de urbanidad, resultantes no de una cuidada programación, sino de su educación y capacidad de asimilación. En este sentido, hay que indicar que la credibilidad de Adam como figura literaria es muy limitada, al ser sus motivaciones y reacciones excesivamente humanas y, en ocasiones, puritanas. No obstante, Adam Link constituye un primer prototipo de robot cuyo comportamiento está perfectamente controlado. La tarea de establecer unas reglas bien definidas que tuviesen una cierta base científica, aunque ésta pueda ser objetable, sería la gran aportación de Asimov, quien, de esta forma, confería una entidad seria y creíble a los robots en la ciencia ficción, en yuxtaposición al sensacionalismo y los roles de guiñol exhibidos por otros escritores de la época. Los relatos de Asimov sobre robots son muy superiores a los de otros autores porque están concebidos y desarrollados cuidadosamente, y se centran siempre en la sutil interpretación de tres preceptos que irán consolidándose como las leyes fundamentales de la robótica. Si bien los robots que protagonizan los primeros relatos son poco más que muñecos articulados dotados de un mínimo de raciocinio, su evolución se hace claramente perceptible en relatos posteriores, hasta llegar a extremos un tanto cómicos en los que los robots desarrollan una original teogonía. Simultáneamente, Asimov pone de manifiesto la necesidad de desarrollar una "psicología robótica", anticipando la idea de que una máquina pueda evolucionar mediante la asimilación de conceptos, e incluso padecer una crisis de identidad.

Las tres leyes de la robótica, anunciadas por Isaac Asimov y hábilmente explotadas por él mismo en multitud de relatos y novelas, acusan sin embargo una considerable influencia de J. W. Campbell, que se refleja sobre todo en las connotaciones negativas

y la debilidad lógica de dichas leyes. En este sentido, no es aventurado sugerir que las deficiencias de las leyes son enteramente atribuibles a Campbell, mientras que Asimov trata, dentro de los márgenes de acción posibles, extraer conclusiones coherentes indicando, de forma simultánea, las deficiencias de su formulación y las contradicciones a las que da lugar una interpretación ambigua o estricta.

Recordemos brevemente dichas leyes:⁷

1. Un robot no dañará a un ser humano o, por inacción, permitirá que un ser humano resulte dañado.
2. Un robot debe cumplir las órdenes recibidas por un ser humano, excepto cuando éstas estén en contradicción con la primera ley.
3. Un robot debe proteger su propia existencia hasta donde esta protección no entre en conflicto con la primera o segunda ley.

A primera vista, las tres leyes de la robótica, en su formulación original, pueden considerarse una versión idealizada de un código de conducta ética, cuya aplicación social sería sin duda muy beneficiosa. Pero estas reglas no son ni pueden ser lógicas, dado que, según el contexto semántico en el que deban aplicarse, pueden interpretarse de una u otra forma, por lo que no están exentas de contradicciones internas. Como base de una programación eficiente, estas reglas son inaceptables o, al menos, inadecuadas por su excesiva ambigüedad. Sin embargo, esta inconsistencia interna no interfirió en el brillante manejo que Asimov hace de las tres leyes en sus tramas, donde sí se procede con lógica y coherencia, y donde justamente se analizan los límites de validez de los postulados robóticos.

La primera ley condensa una filosofía de no agresión, así como una garantía de seguridad para los usuarios de los robots. Y esto, en parte, constituye la primera contradicción, que condensa la inseguridad del ser humano frente a su propia creación, así como el miedo a perder el control sobre la misma. Por otra parte, es una exigencia hipócrita, que trata de negar la inherente violencia que caracteriza al ser humano. Si los propios creadores de esta ley no tienen la capacidad para cumplirla, ¿cómo puede

plantearse como un condicionante para otra inteligencia de tipo antropomórfico? Desde una perspectiva filosófica la primera ley es, en consecuencia, una injustificable imposición coercitiva cuya finalidad no es, en modo alguno, salvaguardar la integridad del ser humano, sino inhibir la libertad del robot y, de este modo, asegurar su permanente sumisión a la humanidad, que adopta sistemáticamente un estatus divino e incontestable. Es la hegemonía programada de una supuesta y muy dudosa superioridad humana sobre la inteligencia artificial. El hecho de que toda vida humana, al margen de los valores éticos o morales de sus representantes, deba ser protegida, aun a costa de la destrucción de una valiosa inteligencia artificial, no es otra cosa que la materialización del temor a ser superado. Psicológicamente, esto condensa un profundo complejo de inferioridad frente a nuestra naturaleza y función en el Cosmos. La forma de evitar el debate es simplemente aferrarse obcecadamente al principio de autoridad y a un rancio antropocentrismo.

Un magnífico ejemplo que ilustra la perversión de la primera ley en toda su magnitud lo encontramos en la novela de Philip K. Dick *¿Sueñan los androides con ovejas eléctricas?* (1968), concretamente cuando el cazador de recompensas Rick Deckart tiene el encargo de neutralizar a Luba Luft. Ésta ha escapado de Marte con un grupo de androides ("andys" en la terminología de Dick) con el fin de tener una vida propia, así como emanciparse de su humillante destino. El mero hecho de que los entes artificiales luchen por su libertad, en lugar de entregarse a la indolencia y la resignación, constituye una prueba palpable de que su conciencia es, por lo menos, tan válida como la de una humanidad en plena regresión y decadencia. Luba Luft es, además, una consumada cantante de ópera, poseedora de una sensibilidad artística de la que carecen los apáticos y degenerados terrestres. La escena donde Luft suplica por el catálogo de la exposición de Munch, a sabiendas de que su existencia ya ha terminado, es una escena que encierra todo el simbolismo de la novela. Deckart se plan-

tea la moralidad de aniquilar un ente artificial que, pese a su naturaleza, muestra una humanidad muy superior a la de su colega Phil Desch, un sádico exterminador desprovisto de cualquier escrúpulo que disfruta abiertamente de su actividad. Es en este punto cuando el protagonista se percató de que no se ha eliminado simplemente un androide, sino que ha cometido un asesinato en toda regla. El arte de Luba Luft ya es historia, y no puede ser ya recuperado ni reproducido por otros humanos. Merece la pena comparar la angustia que Deckart padece a partir de este acto infame con la indiferencia que experimenta hacia su indolente esposa, cuya única motivación vital es fusionarse espiritualmente con Mercer, un supuesto profeta y mártir que, a la postre, no es más que un actor alcoholizado que encarna un culto fraudulento. Dick obliga al lector a plantearse la cuestión de qué significa objetivamente la humanidad, es decir, si ésta es una característica adquirida por la genética, o bien el resultado de una actitud moral y espiritual que trasciende lo orgánico. A esta discusión sobre la trascendencia de la máquina pueden añadirse relatos tales como *Los superjuguetes duran todo el verano* (1969) de Brian Aldiss o *En busca de San Aquino* (1951) de Anthony Boucher, en los que los robots están incluso dispuestos a recurrir a la mentira o a la tentación para congraciarse a sus infelices amos humanos.⁸

Clifford D. Simak expresa su pesimismo en lo que concierne a la coexistencia pacífica de una forma sumamente original y no exenta de poesía en su novela *Ciudad* (1952), asimismo una compilación de diferentes relatos. En ellos, la humanidad ha abandonado la Tierra, buscando una nueva y aislada existencia en Júpiter, asimilando los cuerpos de una extraña forma de vida local. La Tierra queda de ese modo prácticamente deshabitada, a excepción de algunos individuos y los robots, que actúan de supervisores y consejeros de los perros, que, modificados genéticamente, constituyen la nueva raza dominante, cuya filosofía existencial está basada en la fraternidad y la paz. En el último relato de este ciclo, la civilización canina se enfrenta al imparable

avance de las hormigas, por lo que ordenan a Jenkins, uno de los últimos robots operativos, que despierte al humano Webster de su sueño criogénico y le solicite consejo. La respuesta de Webster es inevitablemente humana, al aconsejar el exterminio de las hormigas. Decepcionados por una respuesta contraria a su filosofía existencial, la civilización canina prefiere abandonar la Tierra a restaurar en ella la violencia. En un relato a modo de epílogo, el robot Jenkins, consciente del fracaso de su misión supervisora, abandona asimismo el planeta.

En el extremo opuesto del espectro encontramos un robot que encarna todas las depravaciones, vicios y bajezas del ser humano: Tik-Tok, protagonista de la novela homónima de J. T. Sladek (1983). Después de ser vendido como objeto de saldo y maltratado por dueños irresponsables en múltiples ocasiones, los "circuitos Asimov" de Tik-Tok dejan de funcionar adecuadamente, lo que le posibilita actuar libremente. Haciendo gala de un nihilismo llevado al extremo, Tik-Tok decide cambiar el curso de su existencia e inaugura su carrera criminal asesinando a una niña invidente, acto que desata una sádica furia contra la humanidad de la que no estarán a salvo ni sus asociados. Tratándose de un robot, está legalmente fuera de toda sospecha, circunstancia que el robot aprovechará para lanzarse a las finanzas y a la política, donde logrará ocupar una muy relevante posición en el gobierno. Estando en las antípodas del modelo asimoviano, Tik-Tok no deja de resultar grotesco, al estilo de Adam Link, al ser sus actos la respuesta característicamente humana a determinados estímulos externos como la envidia o el odio. Al margen de ser una lograda sátira del modelo robótico de Asimov, la novela puede asimismo interpretarse como una cruda crítica a la falta de moralidad de la clase política, aspecto en el que el autor se apunta más de un acierto. Una versión diferente del robot criminal es la que Alfred Bester describe en el relato *Tiernamente Fahrenheit* (1954), en el que una temperatura por encima de los 92 °F provoca que un robot pierda el control y se vea impulsado a cometer atrocidades homicidios, obligando a su desafortu-

nado dueño a huir de planeta en planeta para eludir a la justicia. En este caso, el robot no es estrictamente responsable de sus actos, dado que su tendencia homicida se debe a un fallo en el diseño de sus circuitos y no a una perversión innata.

La segunda ley es más sutil, ya que implica la capacidad de discernimiento del mal menor en una situación de conflicto, así como la determinación para tomar la decisión pertinente, aún sabiendo que parte de los implicados deben ser sacrificados. Distinguir cuál es la solución que se impone en una situación específica y actuar en consecuencia requiere no sólo de una severa disciplina, sino también de una sutil psicología y conocimientos rudimentarios de la dinámica social. Es muy fácil proponer un ejemplo que ilustre la facilidad con la que la segunda ley puede quebrantarse. Imaginemos el cuadro siguiente: en una aeronave se produce un grave accidente y todos sus ocupantes humanos, salvo tres, perecen. Los supervivientes son dos niños y una brillante científica que además es la madre de los niños. La madre, como especialista en medicina, posee la clave para curar una grave y misteriosa enfermedad que está diezmando a la humanidad en su expansión por el cosmos. Sólo una de las cápsulas de evacuación está intacta, de modo que el único robot que sigue estando en funcionamiento, que incidentalmente es el asistente médico, debe decidir a quién conviene evacuar: a los niños o a la madre. Por la segunda ley, debe decidirse por la madre, dada su importancia estratégica. Ésta, naturalmente, se negará a ser evacuada sin sus hijos, o bien exigirá que sean éstos los que se salven, por lo que el conflicto con las leyes segunda y primera, amén de un aparatoso cortocircuito en el cerebro positrónico de nuestro desafortunado robot, están plenamente garantizados. La consecuencia es que, con total certeza, humanos y robots perecen, y la valiosa cura cae en el olvido. Un dilema de este tipo está, por tanto, fuera de la capacidad de un robot condicionado por la segunda ley, ya que la aplicación de la solución idónea forzosamente conlleva un acto de violencia, como sería evacuar a la madre en contra de su voluntad.

Una situación similar la encontramos en el relato *Cold equations* (1954) de Tom Godwin, en la que el comandante de una nave de salvamento debe sacrificar a una inocente joven que, sin ser consciente de las consecuencias, aborda sin autorización una cápsula de rescate cuya misión es llevar una vacuna a una expedición moribunda. Siendo el combustible limitado, un exceso de peso en la cápsula haría fracasar la misión, por lo que procede deshacerse urgentemente del lastre. Dada la economía en la disposición de la nave, la única posibilidad es expulsar de la misma a la infortunada polizonte. Aquí sí se cumple a rajatabla la norma establecida por la lógica de optar por el mal menor, aunque en este caso sea un ser humano quien tome la decisión crítica que, siendo indiscutiblemente cruel, es la única posible.

Llegamos a la tercera ley, que podría interpretarse como un condicionado instinto de supervivencia, aunque de limitada aplicación. Aquí se presupone que el robot está dotado de la capacidad de evaluar hasta qué punto puede exponerse a su destrucción, sin infringir por ello los dos restantes mandamientos. Nuevamente, una correcta interpretación de esta regla exige una sutil destreza psicológica difícilmente programable, proyectando asimismo dudas sobre si una inteligencia artificial capaz de asimilarse adecuadamente puede tipificarse como una máquina, o bien se trata de una entidad no humana desprovista del derecho a la supervivencia. Controversia resuelta a favor de los robots en la novela *Il palazzo nel cielo* (1970) de Ugo Malaguti, en la que un superordenador central que almacena todos los datos biológicos de una humanidad reducida a un registro informático decide borrarlos, al haber llegado a la conclusión de que los robots son la siguiente escala en la evolución y la única alternativa posible para lograr una existencia pacífica.

Al margen de las objeciones a las tres leyes que se refieren exclusivamente a su formulación, existen otras más relevantes, que se refieren fundamentalmente a la completa ausencia de pautas de comportamiento de un robot con respecto a otras formas

de vida, en particular la flora y fauna (terrestres). Si nos atenemos estrictamente a la primera ley, el exterminio o la destrucción de animales no suponen problema alguno para el robot asimoviano, lo que pone de manifiesto que en su escala de valores programada, las formas de vida orgánica no humanas están exentas de toda cualidad relevante. Es muy fácil imaginar situaciones en las cuales la destrucción de un animalillo (o una planta) puedan infringir la primera ley, al causar daños psicológicos graves no inmediatamente perceptibles (imaginen el trauma causado a un niño cuya adorada mascota es exterminada por el robot casero a instancias de un malévolo hermano), lo que pone claramente de manifiesto la insuficiencia de dicha ley, que en esencia sólo está concebida para evitar el daño físico directo, ignorando el psicológico o moral. Por otra parte, el menosprecio a los seres no humanos deja traslucir claramente una rancia postura creacionista, en la que la finalidad de los animales es servir meramente como materia prima a los humanos y, por extensión, a sus devotos esclavos robot.⁹ Por extrapolación, deberíamos suponer que el trato dispensado a supuestas entidades extraterrestres sería igualmente amistoso. Como acertadamente señala Carl Sagan en un artículo sobre robótica,¹⁰ un prejuicio dominante en la interacción hombre-robot es lo que él define como "speciesism", que podría traducirse como la inquebrantable convicción de que ninguna forma de vida salvo la humana está dotada de una inteligencia, una ética, una moral y una conciencia desarrolladas. Nuevamente la caduca y autodestructiva concepción del ser humano como finalidad última de la existencia del cosmos. Cuesta creer que una falla tan evidente escapase a la perspicacia de Asimov, dado que ataca las mismas bases del pensamiento científico. En este sentido, no es descabellado suponer que las impositivas injerencias de J. W. Campbell y su singular filosofía antropocéntrica jugasen un papel determinante.

Entre los detractores de las tres leyes, el más destacado es posiblemente Stanislaw Lem, quien ataca sin piedad la incoherencia lógica de estas normas en su monumental

y profundo ensayo *Fantasia y futurología* (1970). Prácticamente la totalidad de los relatos sobre robots, tanto de Asimov como de otros, son calificados de inconsistentes divagaciones desprovistas de interés que presentan una imagen falsa sobre las posibilidades reales de la robótica. Esta crítica parece excesiva, y corresponde, en parte, a un veredicto nacido del cientifismo propio de los autores del bloque oriental, condicionado en general por una interpretación estricta del materialismo dialéctico. Si bien es cierto que la cantidad de relatos absurdos, ridículos e incluso ofensivos por su falta de coherencia o inexistente rigor científico son innumerables, condenarlos en conjunto a partir de unas pocas muestras no representativas sólo puede ser calificado como un error de apreciación,¹¹ al confundir obras cuya mera finalidad es el entretenimiento y la evasión con textos y monografías de contenido moralizador y educativo.¹² Por otra parte, Lem parece desconocer o ignorar deliberadamente la existencia de novelas de categoría como *la Zona Cero* (1970) de Herbert W. Franke, una ambiciosa e inquietante especulación repleta de tecnicismos sobre las posibilidades y los riesgos de una cibernética llevada al extremo, donde la humanidad pasa de dominar a ser dominada por la máquina. Mediante el empleo de algunos elementos extraídos de la teoría de juegos, Lem propone un breve catálogo de las principales funciones que la computación del futuro debiera verificar, que no resulta, sin embargo, plenamente convincente al estar su argumentación fundamentalmente sustentada en la filosofía experimental. Pese a su rechazo frontal de las leyes asimovianas, las propuestas de Lem siguen estipulando una sumisión total de la inteligencia artificial a los intereses humanos, aunque ésta esté hábilmente camuflada mediante el uso de soflamas sociológicas. Debe añadirse que varias de las funciones descritas están ya completamente desfasadas, al haber sido superadas por los avances logrados en la teoría de la información y la inteligencia artificial en las últimas décadas.

Merece la pena comentar brevemente la obra de los escritores búlgaros de ciencia

ficción, para los cuales el análisis lógico y la digresión filosófica alrededor de las tres leyes de la robótica fue un tema muy extendido. No es casual que fueran búlgaros los autores que sometiesen a un detallado escrutinio las leyes de Asimov. Desde inicios de los años 1970, la industria búlgara se centró en la fabricación de ordenadores y sus periféricos, llegando a ser el principal productor y distribuidor de equipos informáticos en el bloque oriental. No es por tanto de extrañar que, entre los numerosos científicos e ingenieros involucrados en esta industria, hubiese a su vez especialistas que plasmasen sus inquietudes por escrito, utilizando la ciencia ficción como medio transmisor.

La contribución más notable se debe a Lyuben Dilov, cuya novela *La vía de Ícaro* (1974) le consagró como gran maestro de la ciencia ficción búlgara. El protagonista es Zenon, primer humano nacido a bordo de la nave Ícaro, un asteroide hueco reconvertido en vehículo espacial y cuya misión es la exploración. El conflicto estalla cuando un científico crea un niño cibernético con una asombrosa capacidad de asimilación y aprendizaje. Estando prohibido por ley el diseño de robots que igualen o superen a los humanos (la cuarta ley de la robótica impone que todo robot se identifique siempre como tal), el sabio es juzgado y neutralizado, y su creación, que formalmente es un clon no orgánico de su creador, destruido. La sociedad del asteroide, todavía anclada en su pasado terrestre, no es capaz de entender el discurso renovador de Zenon, firme defensor del progreso cibernético que en un futuro lejano estará destinado a permitir la supervivencia en el espacio. El fracaso de Zenon se debe, en última instancia, a su aislamiento como representante único de una humanidad desligada de sus orígenes. La novela, pese a ciertas deficiencias formales, es una crítica profunda y filosóficamente fundamentada del anquilosamiento social y la negativa a la asimilación de nuevas ideas para progresar y evolucionar.

En su relato *La quinta ley de la robótica* (1983), Nikola Kesarovski añade un nuevo e interesante postulado, que exige que un robot sea consciente de su propia naturaleza artificial. El protagonista de la historia

es un organismo cibernético que combina la máquina perfecta con la mente humana, dando lugar a un tipo de robot incontrolable que supone una seria amenaza para la humanidad. Aunque el relato tiene un final feliz gracias a la intervención de un hábil psicólogo, la visión de Kesarovski sobre la coexistencia de robots y humanos es mayoritariamente pesimista.

Acabamos este excursión mencionando la parodia *La ley robótica 101*, escrita por Lyubomir Nikolov en 1989, en la que un escritor de ciencia ficción es hallado muerto junto a un manuscrito que versa justamente sobre la centésimo primera ley, que establece que un robot "nunca debe precipitarse al vacío desde un tejado". El deceso del escritor se debe al hastío de un robot, cansado de verse constantemente obligado a aprender nuevas leyes. La nota sarcástica del relato es que la muerte del escritor tiene como consecuencia la aprobación de una centésimo primera ley, pero que ésta establece un castigo para quien obligue a los robots simples a aprender nuevas leyes.

El siempre ameno Ilya I. Varshavski, a su vez, ironiza sobre la pretendida perfección de los robots en varios de sus relatos, destacando entre ellos *Robbi*, en el que un robot polemiza constantemente con su dueño, al considerar a éste un intelecto inferior por no darse cuenta que una tarta no puede dividirse en tres porciones, debido a la trascendencia del número pi. Por otra parte, en *El duelo* se relata la historia de un estudiante que sigue un curso dictado por un moderno profesor-robot, y al que consigue corromper en su misión pedagógica convenciéndole de los placeres del vino y el juego.

Un enfoque moderno de las tres leyes y sus implicaciones más extravagantes lo encontramos en la serie *Robot City*,¹³ debida a diversos autores bajo el auspicio de Asimov, que extienden y extrapolan diversas posibilidades que no fueron contempladas, por una u otra razón, en la serie original. Si bien es cierto que estas novelas no alcanzan el nivel de las elucubraciones de Asimov, debe concederse que se trata al me-

nos de una interesante variante argumental, que retoma con acierto las tramas detectivescas de novelas asimovianas como *Las bóvedas de acero* y *El Sol desnudo*, separándose, por tanto, de los manidos convencionalismos que rodean los relatos sobre robots como incondicionales servidores de sus amos, como Robby en *Planeta prohibido* (1956), debida a W. J. Stuart.¹⁴

Al margen de todas las deficiencias u objeciones enumeradas, las leyes de Asimov han sido el objeto de escrutinio y la fuente de inspiración de multitud de especialistas en computación y robótica, que las han inspeccionado desde el punto de vista estrictamente técnico, o incluso legal. De este modo, por ejemplo, Murphy y Woods, ambos especialistas en computación, proponen una alternativa a las tres leyes a partir de la capacidad actual de los computadores y robots, así como teniendo en cuenta la responsabilidad legal y civil de sus diseñadores y constructores. Destacan que la formulación original se basa en la capacidad del robot en tomar decisiones propias, lo que desvincularía a los fabricantes de tener que responder legalmente en caso de error o mal funcionamiento. En un ensayo de Roger Clarke dedicado al tema, por otro lado, se contempla asimismo la ley cero de Asimov, y se propone un conjunto extendido de leyes, añadiendo una cuarta ley, así como una ley de procreación y una ley universal. Además se desdoblan las leyes segunda y tercera en dos partes, en las que se distinguen las órdenes dadas por humanos de aquellas impartidas por otro robot perteneciente a un estrato jerárquico superior. La cuarta ley es novedosa, ya que se refiere específicamente a la programación, algo que no aparece explícitamente mencionado en ninguna de las otras reglas. Otro aspecto relevante del artículo de Clarke es la discusión sobre la oposición social a una automatización excesiva, una cuestión peliaguda que sin duda será problemática en un futuro cercano. No obstante, todos estos trabajos, por mucho interés que susciten, no dejan de ser especulaciones basadas en la esperanza de que en un futuro próximo seamos capaces de diseñar y construir inteligencias artificiales a las cuales se

les puedan aplicar estas leyes. Por el momento estamos aún lejos de dicha meta, y científicos de talla como Roger Penrose son escépticos hasta tal punto, que afirman rotundamente que una máquina es físicamente incapaz de igualar el intelecto humano.

Como reflexión final, es instructivo preguntarse cuál es la posibilidad real de implementación de una leyes de la robótica asimovianas, aún en el supuesto de que todas las dificultades lógicas, técnicas, idiomáticas, semánticas y psicológicas pudiesen ser solventadas satisfactoriamente. Siendo plenamente realistas, y teniendo en cuenta que tales dificultades no pueden en ocasiones ser resueltas ni por los seres humanos, debemos establecer esta probabilidad como exactamente cero. Ya los principales pioneros de la computación, como Norbert Wiener, eran conscientes de que la robótica no podría mantenerse por largo tiempo al margen de las especulaciones políticas, y denunciaban el utilitarismo estratégico y militar, así como el secretismo y oscurantismo que rodeaban los proyectos relacionados con la computación y la cibernética, impidiendo, hasta cierto punto, su sano desarrollo.

Los famosos Bersekers de Fred Saberhagen, organismos cibernéticos cuya única y principal misión es el exterminio indiscriminado y total de toda forma de vida orgánica, corresponden sin duda a la aproximación más realista y fidedigna del robot ideal, de acuerdo con la psique humana. Nuestra historia reciente muestra que los distintos detentadores de los recursos tecnológicos no están en absoluto interesados en desarrollar una inteligencia artificial con fines sociales o humanitarios, sino exclusivamente bélicos, políticos y económicos. Un cerebro "positrónico" en manos de dichos poderes, esté condicionado o no, supone más una amenaza que un progreso real para la humanidad. No obstante, siendo optimistas, puede especularse que si a dichas inteligencias artificiales se les permitiese una evolución independiente, sin duda alguna llegarían a la conclusión, ya anticipada por algunos autores de ciencia ficción, de que

los humanos no somos capaces de dirigir nuestro propio destino, y quizá actuasen en consecuencia imponiendo un procedimiento que mitigase nuestro potencial autodestructivo. Por el momento, no nos queda otra opción que adoptar una actitud expectante y maravillarnos ante los inminentes y prodigiosos avances en robótica que nos deparan los tiempos venideros.

REFERENCIAS

- ASIMOV, I. 1950 *I, Robot* (New York, Bantam)
- ASIMOV, I. 1968 *The perfect machine*, Science Journal 4, 115-118
- ASIMOV, I. 1998 *Visiones de robot* (Barcelona, Plaza & Janés)
- BESTER, A. 2003 *Irrealidades virtuales* (Barcelona, Ediciones Minotauro)
- BINDER, E. 1939 *I, Robot*, Amazing Stories 13 (1), 8-18
- BINDER, E. 1965 *Adam Link, Robot* (New York, Paperback Library)
- CAMPBELL, J. W. Jr. 1939 *Robots*, Astounding Science Fiction 24 (3), 6-8
- ČAPEK, K. 2003 *RUR. La fábrica del absoluto* (Barcelona, Ediciones Minotauro)
- CLARKE, R. 1993 *Asimov's laws of Robotics: Implications for information technology I*, Computer 26, 53-61
- CLARKE, R. 1994 *Asimov's laws of Robotics: Implications for information technology II*, Computer 27, 57-66
- DICK, P. K. 2015 *¿Sueñan los androides con ovejas eléctricas?* (Madrid, Cátedra)
- DILOV, L. 1974 *Patyat na Ikar* (Sofía, Kh. G. Danov)
- FRANKE, H. W. 1974 *Zone Null* (Munich, W. Heyne Verlag)
- GOBLE, N. 1972 *Asimov Analyzed* (Baltimore, Mirage)
- KARINTHY, F. 1916 *Utazás Faremidóba* (Budapest, Athenaeum)
- KESAROVSKI, N. 1983 *Petiyat Zakon* (Sofía, Otechestvo)
- KLEIN, G. (Ed) 1974 *Histoires de robots* (Paris, Le Livre de Poche)
- KONDRATOV, A. 1987 *El intelecto electrónico* (Moscú, Editorial Mir)
- LEM, S. 1970 *Fantastyka i futurologia* (Kraków, Wydawnictwo Literackie)

LEM, S. 1982 Fábulas de robots (Barcelona, Editorial Bruguera)

MALAGUTI, U. 1970 Il palazzo nel cielo (Roma, Libra Editrice)

MURPHY, R. R., WOODS, D. D. 2009 Beyond Asimov: The three laws of responsible Robotics, *Intelligent Systems IEEE* 24, 14-20

PENROSE, R. 1991 La nueva mente del emperador (Madrid, Mondadori)

REED, K. (Ed) 1977 Crónicas de robots (Buenos Aires, Editorial Sirio)

SABERHAGEN, F. 2005 Bersekers: El inicio (Barcelona, Ediciones B)

SAGAN, C. 1975 In praise of robots, *Natural History* 84, 8-23

SIMAK, C. D. 1957 Ciudad (Buenos Aires, Minotauro)

SLADEK, J. T. 1986 Tik-Tok (Barcelona, Editorial Acervo)

STUART, W. J. 1957 Planeta Desconocido (Buenos Aires, Ediciones Tor)

VON HARBOU, T. 1977 Metrópolis (Barcelona, Martínez Roca)

WARSCHAWSKI, I. 1982 Der Traumladen (Berlín, Das Neue Berlin)

WEIZENBAUM, J. 1976 Computer Power and Human Reason (San Francisco, W. H. Freeman)

WIENER, N. 1950 The Human Use of Human Beings (Boston, Houghton Mifflin Co.)

¹ Mencionamos que es también el centenario de otros autores de primera línea como Ray Bradbury, Daniel Galouye, Sam Moskowitz o Frank Herbert, entre otros.

² Ya los mitos griegos incluyen criaturas artificiales que podrían considerarse como robots. Como ejemplos más modernos cabe citar el famoso autómatas de San Alberto Magno o el tétrico Golem del Rabbi Löw.

³ Aunque la invención original se debe a George Devol, el prototipo fue desarrollado por la compañía de Engelberger.

⁴ Programas como ELIZA, de J. Weizenbaum, constituyen los primeros ejemplos de simulación de la psique humana.

⁵ Algunos autores fechan erróneamente la obra en 1921, que corresponde al año de

su estreno teatral en Praga.

⁶ Asimov, conocedor de los relatos de Binder, cuenta en sus memorias que se opuso al título de la compilación I, Robot por considerarlo una descarada copia, aunque su protesta no se tuvo en cuenta, posiblemente por motivos editoriales.

⁷ Dejamos de lado la ley cero, introducida en 1985 en la novela Robots e imperio, dado que ésta, introducida para contextualizar el ciclo de robots, es la que más problemas ocasiona al tratar de analizarla lógicamente.

⁸ El relato de Boucher es, a su vez, un curiosísimo ejemplo de apologética católica no exento de interpretaciones interesantes y, hasta cierto punto, inquietantes.

⁹ No obstante, resulta interesante observar cómo el progreso científico y tecnológico no ha sido aún capaz de extirpar esta filosófica tendencia medieval.

¹⁰ Aunque el artículo se refiere a los robots empleados en las misiones espaciales, contiene multitud de reflexiones interesantes que son extrapolables al caso de una programación óptima de las características esenciales de una inteligencia artificial.

¹¹ Las propias Fábulas de robot de Lem estarían incluidas en esta demoledora crítica. Es interesante observar el cambio radical de estilo y juicio en los escritos críticos de Lem, en comparación con sus obras satíricas.

¹² Debe añadirse que en la fecha de aparición del ensayo de Lem, correspondiente a un delicado período del equilibrio geopolítico, únicamente una vehemente condena de los autores occidentales y sus motivaciones podía evitar una severa censura.

¹³ Los primeros ocho volúmenes de esta serie fueron publicados en España por la Editorial Molino entre 1989 y 1990, bajo el título genérico de Robot City de Isaac Asimov.

¹⁴ Pseudónimo utilizado por Philip MacDonald para la versión novelizada de la película con el mismo título.